

D. Johnson
#3 9-19-00

Priority papers

1c714 U.S. PTO
09/578371
08/25/00

대한민국 특허청

KOREAN INDUSTRIAL PROPERTY OFFICE

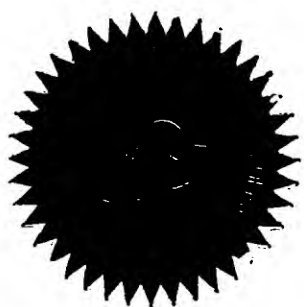
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Industrial
Property Office.

출원 번호 : 특허출원 1999년 제 36905 호
Application Number

출원 년 월 일 : 1999년 09월 01일
Date of Application

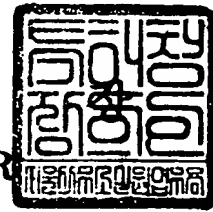
출원 인 : 한국과학기술원
Applicant(s)



2000 년 01 월 28 일

특 허 청

COMMISSIONER



【서류명】	명세서 등 보정서
【수신처】	특허청장
【제출일자】	1999.10.06
【제출인】	
【명칭】	한국과학기술원
【출원인코드】	3-1998-098866-1
【사건과의 관계】	출원인
【대리인】	
【성명】	이원희
【대리인코드】	9-1998-000385-9
【사건의 표시】	
【출원번호】	10-1999-0036905
【출원일자】	1999.09.01
【심사청구일자】	1999.09.01
【발명의 명칭】	외래어 음차표기 유사도 비교 방법
【제출원인】	
【접수번호】	1-1-99-0105458-25
【접수일자】	1999.09.01
【보정할 서류】	명세서등
【보정할 사항】	
【보정대상 항목】	별지와 같음
【보정방법】	별지와 같음
【보정내용】	별지와 같음
【취지】	특허법시행규칙 제13조의 규정에 의하여 위와 같이 제출합니다. 대리인 이원희 (인)
【수수료】	
【보정료】	0 원
【추가심사청구료】	0 원
【기타 수수료】	0 원
【합계】	0 원
【첨부서류】	1. 보정내용을 증명하는 서류_1통

【보정대상항목】 요약

【보정방법】 정정

【보정내용】

본 발명은 외래어에 대한 음차표기 유사도를 비교 판단하는 방법에 관한 것으로 특히, 입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫글자 초성 'ㅇ'을 제외한 모든 초성 'ㅇ'을 제거하는 제 1과정과, 제 1과정을 통해 초성 'ㅇ'이 제거된 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과, 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과, 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과, 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정, 및 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법을 제공하여 외래어 음차표기가 혼재되어 있는 각 자료 혹은 색인어를 혼동없이 빠르게 검색할 수 있으며 한국어에 적합한 유사표기를 검색할 수 있는 기준을 제공한다.

【보정대상항목】 식별번호 10

【보정방법】 정정

【보정내용】

상기 목적을 달성하기 위한 본 발명의 특징은, 입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫글자 초성 'ㅇ'을 제외한 모든 초성 'ㅇ'을 제거하는 제 1과정과, 상기 제 1과정을 통해 초성 'ㅇ'이 제거된 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과, 상기 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과, 상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과, 상기 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정, 및 상기 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 데 있다.

【보정대상항목】 식별번호 24

【보정방법】 정정

【보정내용】

스텝 S101에서는 입력되는 문자를 각 음절마다 자음과 모음 즉, 자소 단위로 풀어쓴 후 첫글자 초성 'ㅇ'(이응)을 제외한 모든 초성 'ㅇ'을 제거한다.

【보정대상항목】 식별번호 25

【보정방법】 정정

【보정내용】

상기 스텝 S101을 통해 초성 'ㅇ'이 제거된 이후, 스텝 S102에서는 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하고, 상기 스텝 S102의 과정을 통해 중복되는 자음을 제거한 후 스텝 S103에서는 첫 글자 초성을 기설정된 대표 자음(표 3참조)으로 변환하게 된다.

【보정대상항목】 식별번호 27

【보정방법】 정정

【보정내용】

【표 2】

자음	코드
ㄱ ㄱ* ㄴ ㅋ	1
ㄴ ㄴ* ㅇ ㅇ*	2
ㄷ ㄷ* ㅌ ㅈ*	3
ㄹ ㄹ*	4
ㄴ ㄴ*	5
ㅂ ㅂ* ㅃ ㅍ ㅎ	6
ㅅ ㅅ* ㅆ ㅈ*	7

【보정대상항목】 식별번호 28

【보정방법】 정정

【보정내용】

【표 3】

자음	대표 자음
ㄱ	ㄱ
ㄴ	ㄴ
ㄷ	ㄷ
ㄹ	ㄹ
ㅁ	ㅁ
ㅂ	ㅂ
ㅅ	ㅅ
ㅇ	ㅇ

【보정대상항목】 식별번호 30

【보정방법】 정정

【보정내용】

상기 표 2에서 '*'가 오른쪽에 붙여진 자음은 종성을 의미한다. Kodex 알고리즘에서는 표준외래어표기법[참조문헌: 문화체육부 고시 제1995-8호 '외래어 표기법' 1995년 3월 16일]에 근거하여 'ㄱ', 'ㄴ', 'ㄷ', 'ㄹ', 'ㅁ', 'ㅂ', 'ㅅ', 'ㅇ' 등의 7개의 종성 자음만을 허용한다. 또한 Kodex 알고리즘에서 사용되는 자음들은 한국어 표준 발음법[참조문헌: 문화교육부 고시 제88-2호 '표준 발음법' 1998년 1월 19일.]에 정의된 19개의 자음들을 사용하였다. 따라서 본 발명에서 사용되는 자음들은 아래의 표 4의 총 26개의 자음들이다.

【보정대상항목】 식별번호 31

【보정방법】 정정

【보정내용】

【표 4】

ㄱ	ㄴ	*	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	*	ㅋ	ㆁ	*	ㄷ	ㄹ	*	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	*	ㅋ	ㆁ	*
ㅂ	ㅅ	*	ㅈ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ
ㅋ	ㆁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ	ㅊ

【보정대상항목】 식별번호 35

【보정방법】 정정

【보정내용】

【표 5】

packet	풀어 쓰기	1단계	2단계	3단계	4단계	5단계	Kodex
패킷	표 ㅂ ㅋ ㅅ *				표 13	표 13	표 13
팩키트	표 ㅂ ㅋ ㅅ —				표 13	표 13	표 13
팩킷	표 ㅂ ㄱ * ㅋ ㅅ *				표 113	표 13	표 13

【보정대상항목】 식별번호 36

【보정방법】 정정

【보정내용】

상기 표 5에서 '패킷', '팩킷', '팩키트'는 모두 본 발명에 따른 외래어 음차표기 유사도 비교 방법을 적용하는 경우에 코드 스트링 '표13'으로 변환되어 일치하므로 같은 영어 단어에서 온 유사한 음차표기라고 판단하는 것이다.

【보정대상항목】 식별번호 37

【보정방법】 정정

【보정내용】

【표 6】

Windows	풀어 쓰기	1단계	2	3	4단계	5	Kodex
윈도우즈	○ T L * C └ ○ T * ─	○ T L C └ T * ─			○ 237		○ 237
윈도우스	○ T L C └ ○ T * ─	○ T L C └ T * ─			○ 237		○ 237
윈도즈	○ T L C └ * ─	○ T L C └ * ─			○ 237		○ 237

【보정대상항목】 식별번호 38

【보정방법】 정정

【보정내용】

【표 7】

sound	풀어 쓰기	1단계	2	3단계	4단계	5	Kodex
싸운드	ㅅㅏ ○ T L * C ─	ㅅㅏ T L C ─		ㅅ T L * C ─	ㅅ 23		ㅅ 23
사운드	ㅅㅏ ○ T L * C ─	ㅅㅏ T L * C ─		ㅅ T L * C ─	ㅅ 23		ㅅ 23

【보정대상항목】 청구항 1

【보정방법】 정정

【보정내용】

입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫 글자 초성 'ㅇ'을 제외한 모든 초성 'ㅇ'을 제거하는 제 1과정과;

상기 제 1과정을 통해 초성 'ㅇ'이 제거된 자소의 나열순서에서 동일자음이 연속 되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과;

상기 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과;

상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과;

상기 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정; 및

상기 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【보정대상항목】 청구항 2

【보정방법】 정정

【보정내용】

제 1항에 있어서,

상기 제 3과정에서의 변환과정 중 기준으로 사용되는 대표 자음은 'ㄱ','ㄷ','ㅂ','ㅅ','ㅈ','ㅊ'인 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【보정대상항목】 청구항 3

【보정방법】 정정

【보정내용】

제 1항에 있어서,

상기 제 4과정에서는 상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음중 'ㄱ, ㄱ*, ㄴ, ㄴ*'에 대하여 자음코드 '1'을 부여하며, 'ㄷ, ㄷ*, ㅇ, ㅇ*'에 대하여 자음코드 '2'를 부여하고, 'ㄷ, ㄷ*, ㅌ, ㅌ*'에 대하여 자음코드 '3'을 부여

하며, 'ㄹ, ㄹ*'에 대하여 자음코드 '4'를 부여하고, 'ㅁ, ㅁ*'에 대하여 자음코드 '5'를 부여하며, 'ㄴ, ㄴ*, ㅂ, ㅂ, ㅅ, ㅎ'에 대하여 자음코드 '6'을 부여하고, 'ㄷ, ㅅ, ㅈ, ㅈ'에 대하여 자음코드 '7'을 부여하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	1999.09.01
【발명의 명칭】	외래어 음차표기 유사도 비교 방법
【발명의 영문명칭】	The similarity comparative method of foreign language tunning fork transcription
【출원인】	
【명칭】	한국과학기술원
【출원인코드】	3-1998-098866-1
【대리인】	
【성명】	이원희
【대리인코드】	9-1998-000385-9
【포괄위임등록번호】	1999-004558-0
【발명자】	
【성명의 국문표기】	최기선
【성명의 영문표기】	CHOI ,Key-Sun
【주민등록번호】	560109-1024017
【우편번호】	305-701
【주소】	대전광역시 유성구 구성동 373-1 한국과학기술원
【국적】	KR
【발명자】	
【성명의 국문표기】	강병주
【성명의 영문표기】	KANG ,Byung-Ju
【주민등록번호】	620829-1063622
【우편번호】	305-701
【주소】	대전광역시 유성구 구성동 373-1 한국과학기술원
【국적】	KR
【발명자】	
【성명의 국문표기】	이재성
【성명의 영문표기】	LEE ,Jae Sung
【주민등록번호】	601015-1351031

【우편번호】	305-701
【주소】	대전광역시 유성구 구성동 373-1 한국과학기술원
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 이원희 (인)
【수수료】	
【기본출원료】	12 면 29,000 원
【가산출원료】	0 면 0 원
【우선권주장료】	0 건 0 원
【심사청구료】	4 항 237,000 원
【합계】	266,000 원
【감면사유】	정부출연연구기관
【감면후 수수료】	133,000 원
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 외래어에 대한 음차표기 유사도를 비교 판단하는 방법에 관한 것으로 특히, 입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫글자 초성 '??'을 제외한 모든 초성 '??'을 제거하는 제 1과정과, 제 1과정을 통해 초성 '??'이 제거된 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과, 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과, 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과, 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정, 및 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법을 제공하여 외래어 음차표기가 혼재되어 있는 각 자료 혹은 색인어를 혼동없이 빠르게 검색할 수 있으며 한국어에 적합한 유사표기를 검색할 수 있는 기준을 제공한다.

【대표도】

도 1

【명세서】**【발명의 명칭】**

외래어 음차표기 유사도 비교 방법{The similarity comparative method of foreign language a tunning fork transcription }

【도면의 간단한 설명】

도 1은 본 발명에 따른 외래어 음차표기 유사도 비교 동작 순서도

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

- <2> 본 발명은 외래어에 대한 음차표기 유사도를 비교 판단하는 방법에 관한 것으로 특히, 영어권에서 보편적으로 사용되고 있는 영어 발음 유사도 비교 알고리즘을 기초로 다양하게 표현되는 외래어의 음차표기 유사도 비교 방법에 관한 것이다.
- <3> 최근 외국과의 교류가 증대함에 따라서 한국어 문서에 많은 외국어의 음차표기가 사용되고 있다. 대부분의 음차표기는 주로 영어에서 온 고유명사나 전문용어이다. 특히 과학 기술 분야의 경우는 영어 전문용어 대한 한국어 번역이 없어 음차표기를 사용할 수밖에 없는 경우가 허다하다.
- <4> 하지만, 외국어에 대한 음차 표기에 개인차가 심하여 이들 텍스트에 대한 문서검색에 어려움을 주고 있다. 예를 들어 영어 'digital'에 대해 '디지털', '디지탈', '디지틀' 등의 세 가지 음차표기가 혼용되어 사용되고 있다. '디지탈'이 표준표기이지만 실제로는 '디지탈'이 더욱 빈번히 사용되고 있고 '디지틀'도 간혹 사용되고 있다.

- <5> 상술한 바와 같은 음차 표기 혼용에 따른 문제는 다양한 음차 표기를 포함하는 문서들에 대한 문서검색에서 이를 고려하지 않을 경우 실제로 검색되어야 하는 문서가 검색되지 못하는 경우가 많이 발생하게 된다는 것으로. 이를 해결하는 한 가지 방법은 색인 시에 동일한 대상 언어의 단어를 표시하는 다양한 한국어 음차표기를 같은 등가부류(equivalence class)로 묶어서 색인하여 놓았다가 질의 시에 이를 자동으로 확장해 주는 것이다[참조문헌: Jeong, K. S., Kwon, Y. H., and Myaeng, S. H., 'The Effect of a Proper Handling of Foreign and English Words in Retrieving Korean Text', In *Proceedings of the 2nd International Workshop on Information Retrieval with Asian Languages* (IRAL' 97), 1997.].
- <6> 이러한 등가부류를 만들기 위해서는 어떠한 두 음차표기가 같은 외국어에서 온 것인지 알아내는 방법, 즉 음차표기 사이의 유사도 비교 방법이 필요하다.
- <7> 또한, 음차표기 유사도 비교방법은 음차표기(외래어) 데이터 베이스에서 근사 검색(approximate search)을 하는데도 반드시 필요한 기술이다. 예를 들어 외래어로 된 상호명 또는 상품명 검색에 유용하게 사용될 수 있을 것이다.
- <8> 그러나, 한국어가 표음문자이기 때문에 발음과 표기가 일치하여 같은 발음에 대해 표기의 애매성이 그다지 심각하지 않기 때문에 한국어 음차표기 유사도 비교 방법은 지금까지 개발된 적이 없고, 따라서, 외국어의 음차표기의 경우의 검색 및 자료관리를 위한 사용자는 매우 큰 불편을 감수하여야 하는 문제점이 발생되었다.
- 【발명이 이루고자 하는 기술적 과제】
- <9> 상기와 같은 문제점을 해소하기 위한 본 발명의 목적은 영어권에서 보편적으로 사

용되고 있는 영어 발음 유사도 비교 알고리즘을 기초로 다양하게 표현되는 외래어의 음차표기 유사도 비교 방법을 제공하는 데 있다.

【발명의 구성 및 작용】

<10> 상기 목적을 달성하기 위한 본 발명의 특징은, 입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫글자 초성 '??'을 제외한 모든 초성 '??'을 제거하는 제 1과정과, 상기 제 1과정을 통해 초성 '??'이 제거된 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과, 상기 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과, 상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과, 상기 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정, 및 상기 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 데 있다.

<11> 본 발명의 상술한 목적과 여러 가지 장점은 이 기술 분야에 숙련된 사람들에 의해 첨부된 도면을 참조하여 후술되는 발명의 바람직한 실시예로부터 더욱 명확하게 될 것이다.

<12> 우선, 본 발명에서 적용되는 기술적 사상을 간략히 살펴보면, 본 발명에 따른 외래어 음차표기 유사도 비교 방법은 영어의 Soundex 알고리즘에서 기본적인 방법론을 차용하였다.

<13> Soundex 알고리즘[참조문헌: Hall, P. and Dowling, G., 'Approximate string matching', *Computing Surveys*, Vol. 12, No. 4, pp. 381-402, 1980.]은 영어권에서 보편적으로 사용되고 있는 영어 발음 유사도 비교 알고리즘으로써, 주로 이름(인명, 지명 등) 데이터베이스에서 찾고자하는 이름의 철자를 정확히 모를 경우 발음으로 근사 검색을 하는데 효과적으로 사용되고 있다.

<14> 상기 Soundex 알고리즘은 영어 단어의 음성적 유사도를 비교하기 위하여 모음을 제거하고 나머지 자음들에 대해서 비슷한 발음을 내는 자음에 동일한 코드를 부여하여 이 Soundex 코드 스트링이 같으면 유사한 발음이라고 판단하는 방법이다.

<15> 구체적인 Soundex 코드 스트링 생성 방법은 다음과 같다.

<16> (1) 모든 모음을 제거한다.

<17> (2) 자음 중에 'H', 'W', 'Y'와 모든 연속되는 중복 자음을 제거한다.

<18> (3) 첫 글자를 제외한 다음 3글자를 아래의 표 1의 Soundex 코드로 치환한다.

<19> 【표 1】

자음	코드
B F P V	1
C G J K Q S X Z	2
D T	3
L	4
M N	5
R	6

<20> 상기 표 1에 대응하여 생성된 두 단어의 Soundex 코드(최대 4글자)가 일치하면 두 단어는 유사한 발음을 가지는 것으로 판단된다.

<21> 상술한 기존의 Soundex 알고리즘을 한글에 적용시키는데는 영어와 한국어의 음운체

계 및 규칙의 차이에 의해 단순 준용시킬수 없으나, 본 발명에서는 기존의 Soundex 알고리즘을 적용시키는데 있어 한국어의 음운적 특성을 고려하여 한국어 실정에 적합하도록 변형한 것이다.

<22> 본 발명에 따른 외래어 음차표기 유사도 비교 방법은 주로 영어에서 온 음차표기 사이의 유사도를 비교하는 방법이다. 따라서, 상기 외래어 음차표기 유사도 비교 방법은 기본적으로 모음을 고려하지 않고 자음들만의 발음 유사도를 비교한다. 대략적인 방법은 발음이 혼돈되는 자음들에 동일한 코드를 부여하여 이 코드 스트링이 일치하면 동일한 영어에서 온 음차표기라고 판단한다.

<23> 구체적인 외래어 음차표기 유사도 비교 방법은 도 1에 도시되어 있는 바와 같다.

<24> 스텝 S101에서는 입력되는 문자를 각 음절마다 자음과 모음 즉, 자소 단위로 풀어 쓴 후 첫글자 초성 '??'(이응)을 제외한 모든 초성 '??'을 제거한다.

<25> 상기 스텝 S101을 통해 초성 '??'이 제거된 이후, 스텝 S102에서는 자소의 나열순서에서 동일자음이 연속되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하고, 상기 스텝 S102의 과정을 통해 중복되는 자음을 제거한 후 스텝 S103에서는 첫글자 초성을 기설정된 대표 자음(표 3참조)으로 변환하게 된다.

<26> 상기 스텝 S103의 과정을 통해 대표자음으로 첫 글자 초성이 변화되어지면 스텝 S104에서는 나머지 자음에 대하여 기설정된 자음 코드(표 2참조)로 치환하게 된다.

<27>

【표 2】

자음	코드
?? ??* ?? ??	1
?? ??* ?? ??*	2
?? ?? ?? ??* ??	3
?? ??*	4
?? ??*	5
?? ??* ?? ?? ??	6
?? ?? ?? ??	7

<28> 【표 3】

자음	대표 자음
??	??
??	??
??	??
??	??
??	??
??	??

<29> 상기 표 3은 첫 글자 초성 대표 자음을 나타내며, 표 2는 본 발명에 따른 외래어 음차 표기 유사도 비교 방법을 적용시키기 위한 모든 사용 자음의 군과 각 군의 코드값을 나타내고 있다.

<30> 상기 표 2에서 '*'가 오른쪽에 붙여진 자음은 종성을 의미한다. Kodex 알고리즘에서는 표준외래어표기법[참조문헌: 문화체육부 고시 제1995-8호 '외래어 표기법' 1995년 3월 16일]에 근거하여 '??', '??', '??', '??', '??', '??', '??' 등의 7개의 종성 자음만을 허용한다. 또한 Kodex 알고리즘에서 사용되는 자음들은 한국어 표준 발음법[참조문헌: 문화교육부 고시 제88-2호 '표준 발음법' 1998년 1월 19일.]에 정의된 19개의 자음들을 사용하였다. 따라서 본 발명에서 사용되는 자음들은 아래의 표 4의 총 26개의 자음들이다.

<31> 【표 4】

?? ??* ?? ?? ??* ?? ?? ?? ??* ?? ??* ?? ??* ?? ?? ??* ?? ?? ??* ?? ?? ?? ?? ?? ?? ??

<32> 이후, 상기 스텝 S104에 따른 과정을 통해 코드화 된 자음 코드는 스텝 S105의 과정을 통해 연속적으로 중복되는 코드를 제거하게되는데, 상기 스텝 S105에서 중복코드가 제거되어지면 이는 스텝 S106의 과정을 통해 기 설정되어 있는 표준 외래어 음차 표기에 대한 코드화 데이터와 비교하게 된다.

<33> 그에 따라, 상기 스텝 S106에서 비교데이터가 동일하다고 판단되면 스텝 S107로 진행하고, 동일하지 않다고 판단노디면 스텝 S108로 진행하게 된다.

<34> 표 5 내지 표 7은 실제적으로 본 발명에 따른 외래어 음차표기 유사도 비교 방법하는 경우의 실 예이다.

<35> 【표 5】

packet	풀어 쓰기	1단계	2단계	3단계	4단계	5단계	Kodex
패킷	????????*?				??13	??13	??13
팩키트	????????????				??13	??13	??13
팩킷	??????*??????*				??113	??13	??13

<36> 상기 표 5에서 '패킷', '팩킷', '팩키트'는 모두 본 발명에 따른 외래어 음차표기 유사도 비교 방법을 적용하는 경우에 코드 스트링 '??13'으로 변환되어 일치하므로 같은 영어 단어에서 온 유사한 음차표기라고 판단하는 것이다.

<37>

【표 6】

Windows	풀어 쓰기	1단계	2	3	4단계	5	Kodex
윈도우즈	????????*????? ?????蝡	????????????????????			??237		??237
윈도우즈	????????*????? ?????蝡	????????????????????			??237		??237
윈도즈	????????*????? ???	????????????????????			??237		??237

<38> 【표 7】

sound	풀어 쓰기	1단계	2	3단계	4단계	5	Kodex
싸운드	????????*?? ??	????????*?? ??		?? ?????*?? ??	??23		??23
사운드	????????*?? ??	????????*?? ??		?? ?????*?? ??	??23		??23

<39> 이상의 설명에서 본 발명은 특정의 실시예와 관련하여 도시 및 설명하였지만, 특허 청구범위에 의해 나타난 발명의 사상 및 영역으로부터 벗어나지 않는 한도 내에서 다양한 개조 및 변화가 가능하다는 것을 당업계에서 통상의 지식을 가진 자라면 누구나 쉽게 알 수 있을 것이다.

【발명의 효과】

<40> 상기와 같이 동작하는 본 발명에 따른 외래어 음차표기 유사도 비교 방법을 제공하여 외래어 음차표기가 혼재되어 있는 각 자료 혹은 색인어를 혼동없이 빠르게 검색할 수 있으며 한국어에 적합한 유사표기를 검색할 수 있는 기준을 제공한다.

【특허청구범위】**【청구항 1】**

입력되는 문자의 표기데이터를 자음과 모음을 분리하여 자소 단위로 분리한 후 첫 글자 초성 '??'을 제외한 모든 초성 '??'을 제거하는 제 1과정과;

상기 제 1과정을 통해 초성 '??'이 제거된 자소의 나열순서에서 동일자음이 연속 되는 경우 나열 순서에서 빠른 순서에 위치하는 해당 자음을 제거하는 제 2과정과;

상기 제 2과정을 통해 중복되는 자음을 제거한 후 첫 글자 초성을 기설정된 대표 자음으로 변환하는 제 3과정과;

상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음에 대하여 기설정된 자음 코드로 치환하는 제 4과정과;

상기 제 4과정을 통해 코드화 된 자음 코드중 종성-초성 관계에 있는 중복코드에서 종성에 해당하는 코드를 제거하는 제 5과정; 및

상기 제 5과정을 통해 중복코드가 제거되어진 후 기 설정되어 있는 표준 외래어 음차표기에 대한 코드화 데이터와 비교하여 동일한 경우 동일 음차표기인 것으로 판단하는 제 6과정을 포함하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【청구항 2】

제 1항에 있어서,

상기 제 3과정에서의 변환과정 중 기준으로 사용되는 대표 자음은 '??','??', '??', '??', '??', '??'인 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【청구항 3】

제 1항에 있어서,

상기 제 4과정에서는 상기 제 3과정을 통해 대표자음으로 첫 글자 초성을 변화시킨 후 나머지 자음중 '??, ??*, ??, ??'에 대하여 자음코드 '1'을 부여하며, '??, ??*, ??, ??*'에 대하여 자음코드 '2'를 부여하고, '??, ??, ??, ??*, ??'에 대하여 자음코드 '3'을 부여하며, '??, ??*'에 대하여 자음코드 '4'를 부여하고, '??, ??*'에 대하여 자음코드 '5'를 부여하며, '??, ??*, ??, ??, ??'에 대하여 자음코드 '6'을 부여하고, '??, ??, ??, ??'에 대하여 자음코드 '7'을 부여하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【청구항 4】

제 3항에 있어서,

자음의 분류중 '*'는 종성에 해당하는 것을 특징으로 하는 외래어 음차표기 유사도 비교 방법.

【도면】

【도 1】

